



# Instructions

---

The Waste Composition Sample Size Calculator has been developed to help municipalities determine the appropriate number of households and areas (sample size) needed to achieve statistical significance. Users can specify their desired confidence level and margin of error, as well as other factors, enabling municipalities to determine the budget required to meet their needs.

## Assistance is available!

The calculator can be adjusted to provide calculations for sample areas with and without area stratification (separating areas based on income, demographics, etc.) and for individual materials or the entire stream. For help adjusting the calculator for your specific needs, please reach out to the CIF.

## Before Getting Started

- Download and install the latest version of R: <https://cran.r-project.org/>
- Download and install the RStudio integrated development environment (free Desktop version): <https://rstudio.com/products/rstudio/download/>
- Download and save the [Waste Audit Sample Package \(waudit\\_1.0.2.tar.gz\)](#) to your computer.
- Download the [Pre-Calculations Worksheet](#).

## Pre-Calculations

You will need to determine the following variables:

**mu\_gar** - Average generation rates for garbage across all the sample areas

**mu\_rec** - Average generation rates for recycling across all the sample areas

**gar\_sd** – Standard deviation of the generation rates for garbage across all the sample areas

**rec\_sd** - Standard deviation of the generation rates for recycling across all the sample areas

If you have data from a previous waste audit, you can use the data to calculate these variables.

The table below outlines the calculations that are to be made, using data from a four-season study with three sample areas. Further explanations are below. The [Pre-Calculations Worksheet](#)

can help you calculate the variables using your own data, or it also contains some sample data that can be used, if you do not have previous waste audit data.

**Generation Rates (kk/hh/wk)**

Sample Area	Garbage				Recycling			
	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter
1	A1	B1	C1	D1	E1	F1	G1	H1
2	A2	B2	C2	D2	E2	F2	G2	H2
3	A3	B3	C3	D3	E3	F3	G3	H3
<b>Average</b>	mu_garA =average (A1,A2,A3)	mu_garB =average (B1,B2,B3)	mu_garC =average (C1,C2,C3)	mu_garD =average (D1,D2,D3)	mu_recE =average (E1,E2,E3)	mu_recF =average (F1,F2,F3)	mu_recG =average (G1,G2,G3)	mu_recH =average (H1,H2,H3)
<b>Standard Deviation</b>	gar_sdA =stdev.s (A1,A2,A3)	gar_sdB =stdev.s (B1,B2,B3)	gar_sdC =stdev.s (C1,C2,C3)	gar_sdD =stdev.s (D1,D2,D3)	rec_sdE =stdev.s (E1,E2,E3)	rec_sdF =stdev.s (F1,F2,F3)	rec_sdG =stdev.s (G1,G2,G3)	rec_sdH =stdev.s (H1,H2,H3)
<b>CV Area</b>	CV_area_garA =gar_sdA /mu_garA	CV_area_garB =gar_sdB /mu_garB	CV_area_garC =gar_sdC /mu_garC	CV_area_garD =gar_sdD /mu_garD	CV_area_recE =gar_sdE/ mu_garE	CV_area_recF =gar_sdF /mu_garF	CV_area_recG =gar_sdG /mu_garG	CV_area_recH =gar_sdH /mu_garH
<b>CV</b>	CV_area =average(CV_area_garA, CV_area_garB, CV_area_garC, CV_area_garD, CV_area_recE, CV_area_recF, CV_area_recG, CV_area_recH)							

### 1. Calculate mu\_gar - Average generation rates for garbage across all the sample areas

For each sample area in your existing study, calculate the total garbage generation rate on a kilograms per household per week basis (kg/hh/wk). This can be as simple as dividing the total garbage collected during the study by the number of households in the study area and dividing the result by the number of weeks the study spanned.



If multiple seasonal studies were conducted, repeat this process for each seasonal study. If there were four seasonal studies, you will have four mu\_gar variables.

## **2. Calculate mu\_rec - Average generation rates for recycling across all sample areas**

For each sample area in your existing study, calculate the total recycling generation rate on a kilograms per household per week basis (kg/hh/wk). This can be as simple as dividing the total recycling collected during the study by the number of households in the study area and dividing the result by the number of weeks the study spanned.

If multiple seasonal studies were conducted, repeat this process for each seasonal study. If there were four seasonal studies, you will have four mu\_rec variables.

## **3. Calculate gar\_sd – Standard deviation of the generation rates for garbage across all the sample areas**

Using the garbage generation rates calculated per sample area, calculate the standard deviation for the study, using the =STDEV.S formula in excel.

If multiple seasonal studies were conducted, repeat this process for each seasonal study. If there were four seasonal studies, you will have four gar\_sd variables.

## **4. Calculating rec\_sd - Standard deviation of the generation rates for recycling across all the sample areas**

Using the recycling generation rates calculated per sample area, calculate the standard deviation for the study, using the =STDEV.S formula in excel.

If multiple seasonal studies were conducted, repeat this process for each seasonal study. If there were four seasonal studies, you will have four rec\_sd variables.

## **5. Calculate cv\_area\_gar – (Formula: gar\_sd / mu\_gar)**

For each season, divide the gar\_sd by mu\_gar. For example, divide the gar\_sd calculated from the Spring Audit by the mu\_gar calculated from the Spring Audit.

If multiple seasonal studies were conducted, repeat this process for each seasonal study. If there were four seasonal studies, you will have four cv\_area\_gar variables.

## **6. Calculate cv\_area\_rec – (Formula: rec\_sd / mu\_rec)**

For each season, divide the rec\_sd by mu\_rec. For example, divide the rec\_sd calculated from the Spring Audit by the mu\_rec calculated from the Spring Audit.

If multiple seasonal studies were conducted, repeat this process for each seasonal study. If there were four seasonal studies, you will have four cv\_area\_rec variables.

## **7. Calculate cv\_area**

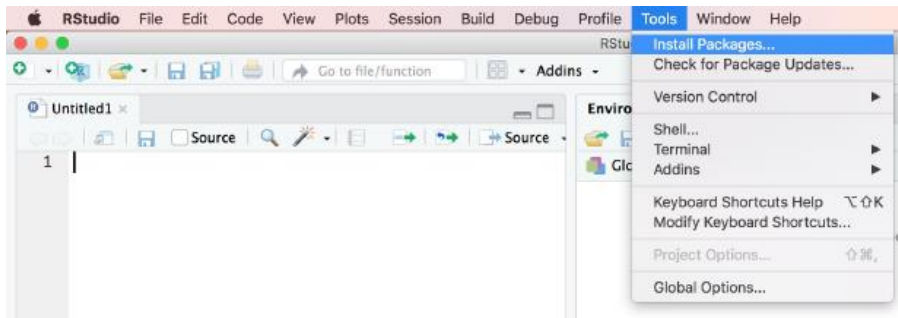
Average all of the values calculated in Step 5 and Step 6 to arrive at a single figure.

## Installing the Program and Package

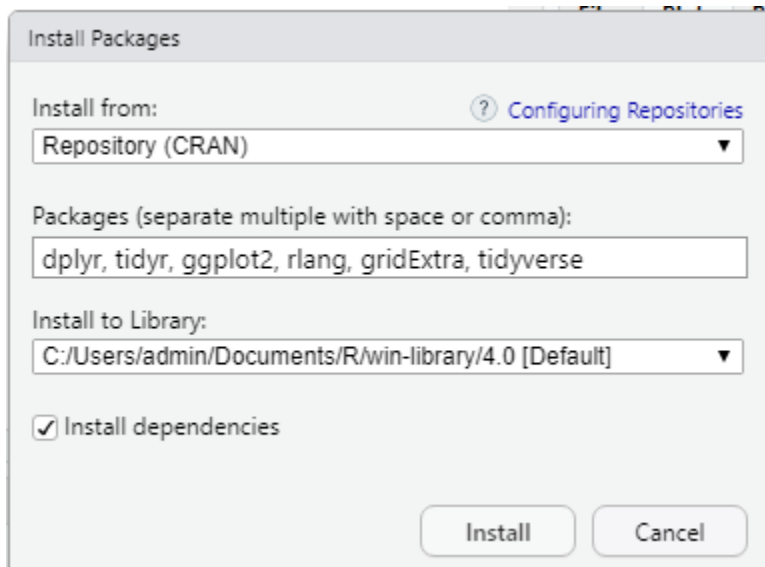
**NOTE:** In all of the code text shown below, **green** is text which should be copied exactly, and **blue** is text which you should input your own values.

### 8. Open RStudio and install the R packages.

Open the RStudio program. As shown below, navigate to **Tools > Install Packages...**



In the dialogue box, type “**dplyr, tidyr, ggplot2, rlang, gridExtra, tidyverse**” to install the required R packages.



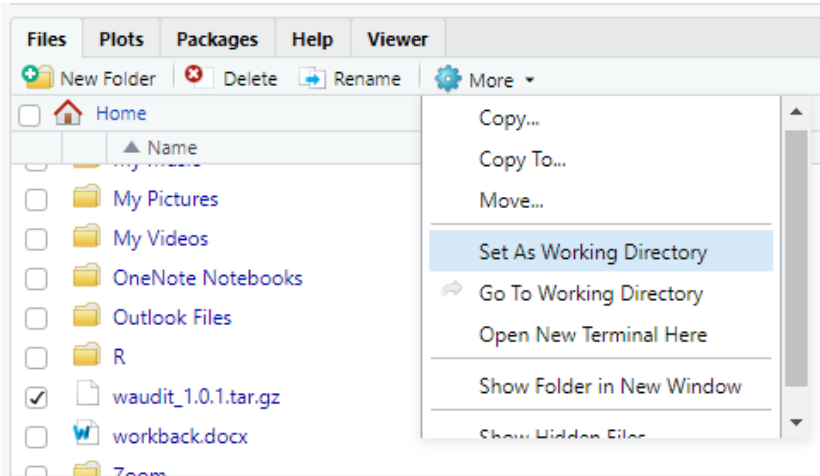
Alternatively, this can be done from the command line (bottom left corner) via:

```
install.packages(c("dplyr", "tidyr", "ggplot2", "rlang", "gridExtra", "tidyverse"))
```



## 9. Install the Waste Audit Sample Package

In the Files window (bottom right corner), navigate to where you saved `waudit_1.0.2.tar.gz`. Click the checkbox beside it, then click **More** and **Set As Working Directory**. This is where R looks for files.



Alternatively, this can be done from the command line (bottom left corner) via:

```
setwd(dir = "path/to/folder/containing/waudit")
```

Next, in the console window (bottom left window), type the command line:

```
install.packages("waudit_1.0.2.tar.gz", repos = NULL, type = "source")
```

Quit and restart R Studio.

In the console window, load the `waudit` package using the command line:

```
require(waudit) # the sample size calculator  
require(tidyverse) # for processing the input file  
require(gridExtra)
```



## Running the Package

NOTE: All of the code provided below needs to be entered within the console section. In all of the code text shown below, green is text which should be copied exactly, and blue is text which you should input your own values.

**10. Enter the four seasonal mu\_gar and mu\_rec values calculated in Steps 1 and 2, by entering the following code:**

```
mu_gar <- c(mu_garA, mu_garB, mu_garC, mu_garD)
mu_rec <- c(mu_recE, mu_recF, mu_recG, mu_recH)
```

**11. Enter the following code to display the value:**

```
# display
rbind(Garbage = mu_gar, Recycling = mu_rec) # true means
```

**12. Enter the cv\_area value calculated in Step 7:**

```
cv_area <- CV_area
```

For example, if the cv\_area calculated in Step 7 was 1.234, then the code to enter would be "cv\_area <- 1.234"

**13. Enter the following code to set-up the calculations:**

```
n_house_data <- 10 # households per area in previous data
n_week_data <- 2 # weeks per household in previous data
```

```
# upper bounds
cv_house <- cv_area * sqrt(n_house_data)
cv_week <- cv_area * sqrt(n_house_data * n_week_data)
```

```
# population inputs
p_inc <- 1 # proportion of population per income level
```

```
# format means as matrices with income levels per row and seasons per column
mu_gar <- matrix(mu_gar, nrow = 1, ncol = 4)
mu_rec <- matrix(mu_rec, nrow = 1, ncol = 4)
```

```
# sampling parameters
n_area <- 10 # number of areas (per income level)
n_house <- 10 # number of households per audit area
n_week <- 2 # number of weeks per household
```



```
# number of simulated audits to generate  
n_sim <- 5e3
```

#### 14. Run the simulator by entering the following code:

```
strat <- simulator_strat(mu_gar = mu_gar, mu_rec = mu_rec, p_inc = p_inc, n_area =  
  n_area, n_house = n_house, n_week = n_week, cv_area = cv_area, cv_house =  
  cv_house, cv_week = cv_week, n_sim = n_sim)
```

#### 15. Display the results from the simulator for garbage and recycling:

```
# garbage amount  
simulator_plot(strat, conf_range = c(.7, .99), output = "gar")  
  
# recycling amount  
simulator_plot(strat, conf_range = c(.7, .99), output = "rec")
```

The result will be a Histogram and a Line Graph. The Histogram represents the average generation rate, while the Line Graph represents the Confidence Level and Margin of Error for the data.

**NOTE:** The Margin of Error is not represented as a percent but rather as kg/hh/wk.

#### 16. Enter the number of areas in the original study:

```
n_unit <- n_area
```

#### 17. Calculate the standard deviations for the data entered.

```
# standard deviation input obtained from the simulator  
gar_sd <- sd(strat$gar) # garbage amount  
rec_sd <- sd(strat$rec) # recycling amount
```

#### 18. Enter the desired Confidence Level using a range of 0 to 1. Example below shows the value of 90% of Garbage and 85% for Recycling:

```
# confidence level for each  
gar_conf <- .9  
rec_conf <- .85
```

#### 19. Enter the desired Margin of Error.

This should be based on the average generation that was calculated in Step 1 of the pre-calculations. For example, if the average generation rate was 5 kg/hh/wk, then a Margin of Error value of 1 kg/hh/wk would effectively represent a 20% Margin of Error.



```
# margin of error for each  
gar_merr <- 1 # units are kg/household/week  
rec_merr <- 0.5 # units are kg/household/week
```

## 20. Run the final code to determine the number of sample areas required:

```
# preliminary sample size estimates for each  
# garbage amount  
gar_size <- sample_approx(merr = gar_merr, alpha = gar_conf, sd = gar_sd, n_unit =  
  n_unit)  
# recycling amount  
rec_size <- sample_approx(merr = rec_merr, alpha = rec_conf, sd = rec_sd, n_unit = n_unit)  
  
# display (rounding sample size to nearest integer)  
ceiling(c(Garbage = gar_size, Recycling = rec_size))  
  
n_size <- sample_approx(merr = c(rec_merr, gar_merr), alpha = c(rec_conf, gar_conf), sd =  
  c(rec_sd, gar_sd), n_unit = n_unit, type = "combined")  
  
n_size <- ceiling(n_size) # round to nearest integer  
n_size # display
```

The final result “n\_size” provides the number of sample areas required to achieve the desired confidence level and margin of error.

That is, if you end up with a value of 15, for example, you would need to include 15 sample areas, with 10 houses each, in order to achieve a 90% confidence level with 1 kg/hh/wk margin of error for garbage, and 85% confidence and .5 kg/hh/wk margin of error for recycling.

If the resulting n\_size (e.g. 15 sample areas), is not achievable, because of time, budget or other resource constraints, then re-run the simulator from Step 18 onwards, with a lower confidence level and/or higher margin of error.

By running the simulator several times, you can understand the relationship between how much data is collected and how confident you can be that the results are representative of the entire population.





## Appendix A: Glossary

### Sample size

A smaller, manageable number of events or observations required to represent a population.

### Statistical significance

The significance level indicates there's a correlation between the sample data and the true population within the confidence level and margin of error indicate the findings are not random.

### Confidence level

The percentage or probability that the average of the sample data collected is also true for the population. The higher the confidence level the greater the reliability of the sample data within a specified margin of error. For example, if a survey is conducted on whether people recycle or not, the results will often be expressed as 80% of people indicate they recycle and the results have a 95% confidence level within a margin of error of 5%.

### Margin of error

The maximum deviation or range the results from the sample data will vary from the true value of the population.

### Simple random sample

A simple random sample is a sampling technique where events or observations are taken at random and there is an equal opportunity for each event or observation to be found in a population or study area.

### Stratified Sample

In a stratified sample, the population is split into sub-groups and therefore the presences of these sub-groups will determine how we sample. For example, a population can be stratified into household income levels.